Greening Multi-Tenant Data Center Demand Response

Niangjun Chen†, Xiaoqi Ren†, Shaolei Ren‡, Adam Wierman†
†California Institute of Technology, Pasadena, CA, USA, E-mail: {ncchen,xren,adamw} @caltech.edu
‡Florida International University, E-mail: sren@cs.fiu.edu

ABSTRACT

Data centers have become critical resources for emergency demand response (EDR). However, currently, data centers typically participate in EDR by turning on backup (diesel) generators, which are both expensive and environmentally unfriendly. In this paper, we focus on "greening" demand response in multi-tenant data centers by incentivizing tenants' load reduction and reducing on-site diesel generation. Our proposed mechanism, ColoEDR, which is based on parameterized supply function mechanism, provides provably near-optimal efficiency guarantees, both when tenants are price-taking and when they are price-anticipating.

1. INTRODUCTION

Power-hungry data centers have been quickly expanding in both number and scale. The flexible energy usage of data centers makes them promising candidates for *demand response*, which is a crucial tool for improving grid reliability and incorporating renewable energy into the power grid.

Emergency Demand Response (EDR) is the most widely-adopted demand response program in the U.S., representing 87% of demand reduction capabilities across all reliability regions. The U.S. EPA has identified data centers as critical resources for EDR which was attested to by the following example: on July 22, 2011, hundreds of data centers participated in EDR by cutting their electricity usage before a large-scale blackout would have occurred [5].

While data centers are increasingly contributing to EDR, they typically participate by turning on their on-site backup diesel generators, which are neither cost effective nor environmentally friendly. For example, in California (a major data center market), a standby diesel generator often produces 50-60 times more nitrogen oxides (a smog-forming pollutant) compared to a typical power plant for each kWh of electricity, and diesel particulate represents the state's most significant toxic air pollution problem [8].

Consequently, modulating server energy for green EDR has received an increasing amount of attention in recent years. For example, a recent field study by LNBL has demonstrated that data centers can reduce energy consumption by 10-25% for demand response (by, e.g., turning off unused servers and/or migrating workloads to other sites), without noticeably impacting normal operation [2].

While existing studies on data center demand response show promising progress [4, 9], they are primarily focused on owner-operated data centers (e.g., Google) whose operators have full control over both servers and facilities. In this paper, we focus on another type of data centers — multi-tenant colocation data centers (e.g., Equinix). In a colocation data center (simply called "colo"), multiple tenants deploy and keep full control of their own physical servers in a shared space, while the colo operator only provides facility support (e.g., power and cooling). Colos are much less studied than owner-operated data centers, but they are actually very common, consuming nearly 40% of total energy by data centers [6].

In addition, unlike many mega-scale owner-operated data centers built in rural areas, colos are mostly located in metropolitan areas, where EDR is most needed. For all these reasons, colos are key participants in EDR programs.

Greening EDR in multi-tenant data centers, however, relies on reducing server energy from tenants which may not have incentives to do so, thus raising the research question: how can a colo operator *efficiently* incentivize its tenants' load shedding for EDR?[1]

Overview of results. This abstract presents a summary on our proposed mechanism, ColoEDR, based on supply function bidding to "green" colocation demand response by incentivizing load reduction from tenants instead of fully relying on backup diesel generation [1].

To our best knowledge, this paper represents the first attempt to design a supply function bidding mechanism for colocation demand response. Our approach builds on, and also adds to, the supply function literature. Furthermore, we show that ColoEDR suffers little performance loss compared to the socially optimal outcome, both when tenants are price-taking and when they are price-anticipating.

2. PROBLEM FORMULATION

A common type of demand response program is mandatory EDR: participants typically sign contracts with a load serving entity (LSE) in advance (e.g., 3 years ahead in PJM [7]) and receive financial rebates for their committed load reduction even if no EDR signals are triggered, whereas non-compliance (i.e., failure to cut load as required during EDR) incurs heavy penalty [7].

When the operator receives an EDR signal from the LSE, it has two options for satisfying the load reduction. First, without involving the tenants, the colo operator can use the on-site backup diesel generator. We denote the amount of energy reduction using diesel generation by y and the cost per kWh of diesel generation (e.g., for fuels) by α.

Alternatively, the colo operator could try to extract IT energy reductions from the tenants. We consider a setting

[1] Tenants are ineligible to participate in EDR program since they receive UPS-protected power from colo operator and are not directly connected to the grid.

where there are N tenants, $i \in \mathcal{N} = \{1, 2, \cdots, N\}$. When shedding energy consumption, a tenant i will incur some costs and we denote the cost from shedding s_i by a function $c_i(s_i)$. These costs could be due to wear-and-tear, performance degradation, workload shifting, etc. We make a standard assumption that the cost functions are continuous, convex, and strictly increasing.

We introduce a simple and practical mechanism ColoEDR below, and then discuss it in detail in the text that follows.

1. The colo operator receives an EDR reduction target δ and broadcasts to tenants a parameterized supply function $S(\cdot, p)$ according to
$$S(b_n, p) = \delta - \frac{b_n}{p}. \tag{1}$$
where p is offered reward for each kWh of energy reduction and b_n is the bidding values that can be chosen by tenant n.

2. Participating tenants respond by placing their bids b_n;

3. The colo operator decides the amount of on-site generation y and market clearing price p. Given any y, the market clearing price has to satisfy $\Sigma_n S(p(\mathbf{b}), b_n) + y = \delta$,[2] and thus
$$p(\mathbf{b}, y) = \frac{\sum_n b_n}{(N-1)\delta + y}. \tag{2}$$
To determine the amount of local generation y, the operator minimizes the cost of the two load-reduction options, i.e.,
$$y = \underset{0 \le y \le \delta}{\arg\min} (\delta - y) \cdot p(\mathbf{b}, y) + \alpha y. \tag{3}$$

4. EDR is exercised. $\forall n \in \mathcal{N}$, tenant n sheds $S(b_n, p)$, and receives $pS(b_n, p)$ reward.

There are several advantages of ColoEDR from an operation point of view. First, bidding for the tenants is simple – they only need to communicate one number, and it is already common practice for operators to communicate with tenants before EDR events, so the overhead is marginal. Second, the colo operator collects just enough information (i.e., how much energy reduction each tenant will contribute to EDR), while tenants' private cost function is masked by the form of the supply function and hence not solicited. Third, ColoEDR guarantees that the colo operator will not incur a higher cost than the case where only diesel generator is used. Further, ColoEDR pays a uniform price to all participating tenants and hence ensures fairness.

Our mechanism is most related to [3], which considers an inelastic demand δ that must be satisfied via N suppliers using supply function mechanism and proves efficient bounds on the resulting equilibrium. In contrast, our work assumes that the market operator has an outside option (diesel) that can be used to satisfy the inelastic demand. This leads to a multistage game between the tenants and the profit-maximizing operator, a dynamic that has not been studied previously in the supply function literature.

VCG-based mechanisms can be considered as natural alternatives to our approach. While these mechanisms are truthful, they violate all the four properties discussed above. Under such approaches, tenants must submit complex bids revealing their true costs; payment made to tenants may be unbounded, and prices to tenants are differentiated, raising unfairness issues.

[2] For ease of presentation, we assume the power usage efficiency (PUE) to be 1 here. Otherwise, one can simply scale y, α and δ with PUE accordingly.

3. EFFICIENCY ANALYSIS OF ColoEDR

To evaluate the efficiency of ColoEDR, we use the (socially) optimal outcome as a benchmark. This outcome relies on the operator having full knowledge of tenants' costs and full control over tenants' energy reduction, and tries to minimize the social cost by solving the following problem.

$$\begin{aligned} \mathsf{SCM}: \quad & \min \ \alpha y + \sum_{i \in \mathcal{N}} c_i(s_i) && \text{(4a)} \\ \text{s.t.} \quad & y + \sum_{i \in \mathcal{N}} s_i = \delta && \text{(4b)} \\ & s_i \ge 0, \ \forall i \in \mathcal{N}, \quad y \ge 0. && \text{(4c)} \end{aligned}$$

The objective in SCM can be interpreted as the tenants' cost plus the colo operator's cost. Note that the internal payment transfer between the colo operator and tenants cancels, and does not impact the social cost. Also, the Lagrange multiplier of (4b) can be interpreted as the social optimal price p^*, i.e., given this price as reward for energy reduction, each tenant will individually reduce their energy by s_n that corresponds to the social cost minimization solution in (4).

3.1 Price-Taking Tenants

When tenants are price-taking, they maximize their net utility, which is the difference between the payment they receive and the cost of energy reduction, given by:

$$\begin{aligned} P_n(b_n, p) &= pS_n(b_n, p) - c_n(S_n(b_n, p)) \\ &= p\delta - b_n - c_n\left(\delta - \frac{b_n}{p}\right). \end{aligned}$$

Here, the price-taking assumption implies that the variable p is considered to be as is. It normally holds when market consists of many players of similar sizes, who each has little power to impact the market clearing price. The market equilibrium for price-taking tenants is thus defined as follows.

Definition 1. *A triple $(\mathbf{b}, p, y)$ is a (price-taking) market equilibrium if $P_n(b_n; p) \ge P_n(\bar{b}_n; p) \quad \forall n \in \mathcal{N}, \bar{b}_n \ge 0$, and market is cleared by setting the price p according to (2), and the amount of on-site generation is decided by (3).*

The key to our analysis is a complete characterization of the market equilibrium as the solution of an optimization problem below. This characterization leads to efficiency bounds for ColoEDR in Table 1.

Theorem 1. *When tenants are price-taking, there is a unique market equilibrium $(\mathbf{b}^t, p^t, y^t)$ for ColoEDR, and the resulting allocation $(\mathbf{s}^t, y^t)$ can be characterized by the optimal solution of the following problem*

$$\begin{aligned} \min_{\mathbf{s}, y} \quad & \sum_n c_n(s_n) + \frac{\alpha}{2N\delta} \left(y + (N-1)\delta\right)^2 \\ s.t. \quad & \sum_n s_n = \delta - y, \\ & s_n \ge 0, \ \forall n, \quad y \ge 0. \end{aligned}$$

3.2 Price-Anticipating Tenants

In contrast to the price-taking model, price-anticipating tenants realize that they can change the market price by their bids, and adjust their bids accordingly. In particular, they realize that the local generation and market price are set by (2) and (3), and by first order optimality,

$$y(\mathbf{b}) = \sqrt{\frac{(\Sigma_n b_n)N\delta}{\alpha}} - (N-1)\delta; \quad p(\mathbf{b}) = \sqrt{\frac{(\Sigma_n b_n)\alpha}{N\delta}}. \quad (5)$$

The price-anticipating model is suitable when the market consists of a few dominant players, who have significant power to impact the market price through their bids, i.e., the oligopoly setting.

Given bids from the other tenants $\mathbf{b}_{-n}$, each price-anticipating tenant n optimizes the following cost over bidding value b_n

$$Q_n(b_n; \mathbf{b}_{-n}) = p(\mathbf{b})S_n(b_n, p) - c_n(S_n(b_n, p))$$

Thus, substituting (1) and (5), we have

$$Q_n(b_n; \mathbf{b}_{-n}) = \sqrt{\frac{(\Sigma_n b_n)\alpha\delta}{N}} - b_n - c_n\left(\delta - \frac{b_n}{\sqrt{\Sigma_m b_m}}\sqrt{\frac{N\delta}{\alpha}}\right).$$

The price-anticipating equilibrium is defined as follows:

Definition 2. *A triple $(\mathbf{b}, p, y)$ is a (price-anticipating) market equilibrium if $Q_n(b_n; \mathbf{b}_{-n}) \geq Q_n(\bar{b}_n; \mathbf{b}_{-n})\ \forall n \in \mathcal{N}$, $\bar{b}_n \geq 0$, the market is cleared by setting the price p according to (2) and the amount of on-site generation is decided by (3).*

Similar to the price-taking case, we again show that there exists a unique equilibrium and, furthermore, that the tenants and operator behave in equilibrium as if they were solving an optimization problem of the same form as the aggregate cost minimization (4), but with "modified" cost functions.

Theorem 2. *When tenants are price-anticipating, there is a unique equilibrium $(\mathbf{b}^a, p^a, y^a)$ for ColoEDR, and the resulting allocation $(\mathbf{s}^a, y^a)$ can be characterized by the optimal solution to the following optimization:*

$$\begin{aligned} \min \quad & \sum_n \hat{c}_n(s_n) + \frac{\alpha}{2N\delta}\left(y + (N-1)\delta\right)^2 \\ s.t. \quad & \sum_n s_n = \delta - y \\ & s_n \geq 0,\ \forall n, \quad y \geq 0, \end{aligned}$$

where, for $s_n \geq 0$,

$$\begin{aligned} \hat{c}_n(s_n) =& \frac{1}{2}\left(c_n(s_n) + s_n\frac{\alpha}{2N}\right) \\ &+ \frac{1}{2}\int_0^{s_n}\sqrt{\left(\frac{\partial^+ c_n(z)}{\partial z} - \frac{\alpha}{2N}\right)^2 + 2\frac{\partial^+ c_n(z)}{\partial z}\frac{z\alpha}{N\delta}}dz, \end{aligned}$$

and for $s_n < 0$, $\quad \hat{c}_n(s_n) = 0$.

Although the form of $\hat{c}_n(s_n)$ looks complicated, there is a simple linear approximation that gives useful intuition.

Lemma 3. *For all modified cost $\hat{c}_n, n \in 1, \ldots, N$, for any $0 \leq s_n \leq \delta$, $c_n(s_n) \leq \hat{c}_n(s_n) \leq c_n(s_n) + s_n\frac{\alpha}{2N}$.*

The form of Lemma 3 shows that the difference between the modified cost function and the true cost diminishes as N increases, and this is the key observation that underlies our subsequent results upper bounding the efficiency loss of ColoEDR.

3.3 Discussion

With the characterization of the equilibria in Theorem 1 and 2, we can now bound the efficiency of ColoEDR. The main results for the price-taking and price-anticipating analyses are summarized in Table 1 and 2, where price ratio compares the market price, colo saving compares operator's cost of EDR, and welfare loss compares the social cost of market equilibrium with the optimal allocation. In Table 2, price markup is the increase in price when tenants are price-anticipating compared to when they are price-taking.

Tenants	Price Ratio	Colo Saving	Welfare Loss
Price-taking	$[\frac{N-1}{N},\ 1]$	$[0,\ \alpha\delta/N]$	$[0,\ \alpha\delta/2N]$
Price-anticipating	$[\frac{N-1}{N},\ 1]$	$[0,\ \alpha\delta/N]$	$[0,\ \alpha\delta/N]$

Table 1: Performance guarantee of ColoEDR compared to the social optimal allocation.

Price Markup	Extra Diesel	Colo Saving
$[0,\ \alpha/2N]$	$[0,\ \delta/2]$	$[-\alpha\delta/N,\ 0]$

Table 2: Comparison of price-anticipating against price-taking equilibria of ColoEDR.

To summarize the results in Table 1 and 2 briefly, note first that ColoEDR always benefits the operator, since the colo saving is always non-negative compared to the diesel-only method. Second, regardless of tenants being price-taking or price-anticipating, ColoEDR is approximately socially cost-minimizing as the number of tenants grows. Third, ColoEDR also gives tenants approximately social optimal payment, since both the operator's additional benefit and welfare loss are bounded by $\alpha\delta/N$. The loss in payment for tenants compared to the social optimum is at most $2\alpha\delta/N$, which approaches 0 as N grows. Lastly, Table 2 shows that while price-anticipating tenants drive up the market price and increase the cost of the operator, ColoEDR is robust against this behavior and the loss approaches 0 as N grows. Further, it also highlights that it is possible for price-anticipating tenants to cause the operator to use more diesel generation, though our case study [1] shows that this is very unlikely to occur in practice.

4. REFERENCES

[1] N. Chen, X. Ren, S. Ren, and A. Wierman. Greening multi-tenant data center demand response. *arXiv preprint arXiv:1504.07308*, 2015.

[2] G. Ghatikar, V. Ganti, N. E. Matson, and M. A. Piette. Demand response opportunities and enabling technologies for data centers: Findings from field studies, 2012.

[3] R. Johari and J. N. Tsitsiklis. Parameterized supply function bidding: Equilibrium and efficiency. *Operations research*, 59(5):1079–1089, 2011.

[4] Z. Liu, A. Wierman, Y. Chen, B. Razon, and N. Chen. Data center demand response: Avoiding the coincident peak via workload shifting and local generation. *Performance Evaluation*, 70(10):770–791, 2013.

[5] A. Misra. Responding before electric emergencies, http://www.afcom.com/digital-library/pub-type/communique/responding-before-electric-emergencies/.

[6] NRDC. Scaling up energy efficiency across the data center industry: Evaluating key drivers and barriers. *Issue Paper*, Aug. 2014.

[7] PJM. Emergency demand response (load management) performance report – 2012/2013. Dec. 2012.

[8] Santa Babara County. Air Pollution Control District, http://www.ourair.org/do-you-really-need-a-diesel-generator/.

[9] A. Wierman, Z. Liu, I. Liu, and H. Mohsenian-Rad. Opportunities and challenges for data center demand response. In *IGCC*, 2014.